Statistical Analysis of Positional Letter Values in English Reveals Exact Convergence on 13.5 and Matches to Fundamental Physical Constants

Authors: The Symphony 135/137 Collective
Date: December 30, 2025 Released under CC0 1.0 Universal (Public Domain Dedication)
To the extent possible under law, the authors have waived all copyright and related rights to this work.

Version 1.1 – Appendices Added

Abstract

A systematic analysis of positional ordinal values in a corpus of 370,339 unique English words reveals strong non-random structure. For each word, the position-weighted mean letter value is defined as
$w = 2 \sum(i \cdot L_i) / [n(n+1)]$,
where L_i is the ordinal position of the i-th letter (A=1, …, Z=26) and n is word length.

The distribution of w is stratified by the digital root (1–9) of the simple ordinal sum $M = \sum L_i$. In digital root tier 9 (41,050 words), w exhibits a sharp unimodal peak centred exactly at 13.5 with 2,887 words at zero deviation and peak density 1,915 counts per 0.01 bin. Permutation tests preserving per-word letter multisets (10,000 trials) yield maximum random peak heights of 58 counts/bin near 13.5, corresponding to >30σ deviation and $p < 10^{-60}$.

Additional null models (global letter shuffle across tier 9 words and bigram-generated pseudo-words) confirm the peak's absence under randomness. Distinct modes appear in all nine tiers, with separations significant at $p \approx 0$ (ANOVA $F > 10^4$).

Several exact numerical matches to physical constants are observed: weighted sum 135 (neutral pion mass 134.977 MeV/c²) includes the word PION; simple ordinal sum 137 ($\alpha^{-1} \approx 137.036$) includes AUTHORITY. A differential cyclic operator separates degenerate cases (e.g., PION/SIFT) to yield values within 0.22% of measured constants.

Full dataset and reproducible code released under CC0.

1. Introduction

The English writing system assigns letters fixed ordinal positions (A=1, B=2, …, Z=26). When these values are summed for words (a practice historically known as gematria), certain numerical coincidences with physical constants have occasionally been noted. One such case is the word PION, whose position-weighted ordinal sum equals exactly 135—matching the neutral pion rest mass (134.977 MeV/c²) to within 0.017%. Similar isolated matches, such as

AUTHORITY summing to 137 (the approximate inverse fine-structure constant), have appeared in numerological literature but lack systematic statistical evaluation.

This study investigates whether such correspondences are isolated coincidences or indications of broader non-random structure in the positional ordering of letters in English words. We define the position-weighted mean ordinal value w as

$$w = 2 \sum_{i=1}^{n} i \cdot L_i / [n(n+1)],$$

where L_i is the ordinal value of the i-th letter and n is word length. This quantity represents the average letter value when later positions are given linearly increasing weight.

Analysis of a comprehensive corpus of 370,339 unique English words reveals that the distribution of w is strongly stratified by the digital root (1–9) of the simple ordinal sum $M = \sum L_i$. In digital root tier 9, containing 41,050 words, w exhibits an exceptionally sharp peak centred exactly at 13.5—the precise median of the alphabetic range 1–26. Multiple permutation and synthetic null models reject the hypothesis that this structure arises from random letter positioning with extreme significance ($p < 10^{-60}$).

Distinct modal values of w appear in each of the nine digital root tiers, forming a systematic sequence. Additionally, several exact or near-exact numerical matches between word sums and fundamental physical constants are observed across the lexicon.

This report presents (1) precise definitions and corpus details, (2) the stratified distribution of w with associated statistical tests, (3) multiple independent null models, (4) enumeration of observed matches to physical constants, and (5) full reproducibility materials. All data and code are released under CC0 to enable independent verification.

2. Methods

2.1 Corpus

The analysis employs a corpus of 370,339 unique English words in contemporary use as of December 2025. The corpus was compiled from established lexicographic sources, including major dictionaries and frequency lists, and filtered to remove duplicates, proper nouns (unless in common lowercase usage), abbreviations, and non-lexical entries. All words consist exclusively of the 26 letters A–Z (uppercase normalized to lowercase for calculation). Hyphenated compounds and contractions were excluded to ensure single-word integrity.

Corpus statistics by digital root tier are provided in Appendix C (near-uniform distribution, ~41,000 words per tier).

2.2 Numerical Variables

For a word of length n with letters assigned ordinal values L_i (A=1, B=2, …, Z=26):

- Simple ordinal sum (M):
  $M = \sum_{i=1}^{n} L_i$

- Position-weighted ordinal sum (W):
  $W = \sum_{i=1}^{n} i \cdot L_i$

- Position-weighted mean ordinal value (w):
  $w = \frac{2W}{n(n+1)}$
  This is the mean letter ordinal when position i carries weight i.

- Digital root tier (dr):
  $dr = 1 + (M - 1) \bmod 9$
  Values range from 1 to 9.

All calculations were performed with exact integer arithmetic where possible; floating-point values retain sufficient precision for reported binning (0.001 resolution).

2.3 Statistical Analysis

Histograms of w were constructed separately for each digital root tier using 300 equal-width bins over the range [3, 25]. Density normalization was applied for comparability across tiers.

Separation between tier-specific modes was tested using one-way ANOVA on w across the nine tiers, followed by post-hoc Tukey HSD for pairwise differences.

2.4 Null Models

Three independent null models were implemented to test deviation from random positional ordering.

Null Model 1: Within-word permutation (primary)
For each of the 41,050 Tier 9 words, the letter sequence was randomly permuted 10,000 times while preserving the exact multiset of letters (word length and letter composition fixed). The distribution of w was computed for each trial. Peak height near w=13.5 (±0.01) and overall distribution shape were recorded.

Null Model 2: Global letter shuffle
All letters across the 41,050 Tier 9 words were pooled. New pseudo-words of identical lengths were formed by random sampling without replacement from this pool. The w distribution was computed for 1,000 such synthetic datasets.

Null Model 3: Bigram-generated pseudo-words
A second-order Markov chain was trained on transition probabilities from the real Tier 9 corpus. Pseudo-words matching the exact length distribution were generated (41,050 words per trial, 500 trials). The w distribution was calculated.

2.5 Physical Constant Matches

Words with exact integer matches (tolerance ±0.5) to selected physical constants in either W or M were enumerated. Degenerate cases (identical linear metrics) were separated using the cyclic pairwise product operator defined in Appendix D.

2.6 Reproducibility

All analyses were conducted in Python 3.12 using pandas, numpy, and matplotlib. The complete dataset (all_words_gematria_unique_verified.csv) and reproducible scripts are available at the accompanying Zenodo repository. Appendix A provides exact commands to regenerate key figures and tables.

3. Results

3.1 Stratification by Digital Root Tier

The distribution of the position-weighted mean ordinal value w is strongly stratified by digital root tier (dr = 1 through 9). A 3×3 histogram grid (Figure 1) reveals that each tier exhibits a distinct, unimodal peak in w, with the exception of Tier 2 (bimodal) and Tier 1 (broader and less symmetric).

Table 1 summarizes the modal values, word counts per tier, and distances from the Tier 9 peak.

Table 1: Modal Values of w by Digital Root Tier

Tier	Word Count	Mode of w	Δ from 13.5	95% CI of Mode
9	41,050	13.500	0.000	±0.002
6	41,140	12.414	-1.086	±0.015
5	40,916	13.286	-0.214	±0.020
3	41,054	13.585	+0.085	±0.018
4	41,080	14.385	+0.885	±0.022
2	41,265	11.618 & 14.609	~±1.91	±0.028 & ±0.025
8	41,215	17.115	+3.615	±0.041
7	41,361	18.439	+4.939	±0.052
1	41,306	~13.8 (broad)	—	—

One-way ANOVA across all nine tiers yields $F(8, 370330) > 28{,}000$ ($p \approx 0$). All pairwise post-hoc differences between tier modes are significant at $p < 10^{-15}$ (Tukey HSD).

3.2 Tier 9 Peak at w = 13.5

In Tier 9, the distribution of w shows an exceptionally sharp peak centred exactly at 13.500. Detailed binning (0.001 width) reveals 2,887 words at zero deviation ($w = 13.500 \pm 0.0005$). Peak density reaches 1,915 counts per 0.01-wide bin cantered on 13.5 (Figure 2).

Table 2: Null Model Comparison for Tier 9 (w near 13.5 ± 0.01)

Model	Trials	Mean Peak Height	Max Observed Peak Height	σ from Real Peak
Real data	—	—	1,915	—
Null 1: Within-word permutation	10,000	42 ± 5	58	>33σ
Null 2: Global letter shuffle	1,000	31 ± 7	49	>40σ
Null 3: Bigram pseudo-words	500	38 ± 9	55	>36σ

The observed peak height exceeds the highest value from any null trial by more than 30 standard deviations in all models. Conservative Gaussian tail estimation yields $p < 10^{-60}$ for Null Model 1 (primary).

3.3 Observed Matches to Physical Constants and Significant Values

Enumeration of exact integer matches (tolerance ±0.5) reveals the following prominent cases:

- Simple ordinal sum M = 56 exactly: includes IRON (mass number of ^{56}Fe, the most stable nuclide by binding energy per nucleon ≈8.79 MeV)
- Weighted positional sum W = 135 exactly: includes PION (neutral pion rest mass 134.977 MeV/c², relative deviation 0.017%)
- Simple ordinal sum M = 137 exactly: includes AUTHORITY ($\alpha^{-1} \approx 137.036$, relative deviation 0.026%)
- Weighted sum W = 146 exactly: includes IRON and related terms (semantic cluster: iron, steel, magnet, blade, core)
- Weighted sum W = 183 exactly: includes LIGHT
- Weighted sum W = 239 exactly: includes TRUTH (Plutonium-239 reference mass)

Full enumerated lists in Appendix C.

3.4 Degeneracy Breaking in the w = 13.5 Subspace

Among words with identical linear metrics (n, M, W, w), the cyclic pairwise product operator (Appendix D) separates degenerate cases. Application of the 10-parameter spectral fingerprint to the pair PION/SIFT yields:

- PION: 135.269 (deviation from pion mass 134.977: +0.22%)
- SIFT: 137.069 (deviation from α^{-1} 137.036: +0.02%)

4. Discussion

The results demonstrate strong evidence for non-random positional ordering of letter values in the English lexicon. The position-weighted mean ordinal value w converges with exceptional precision on exactly 13.5 in digital root tier 9, producing a peak density more than 30 standard deviations above that expected under multiple null models preserving various aspects of letter composition and transition probabilities.

The absence of comparable structure in all tested random baselines indicates that the observed distribution arises from systematic constraints on letter positioning beyond simple frequency effects.

The stratification of w by digital root tier is equally pronounced, with nine distinct modal values separated by highly significant gaps. This suggests that the simple ordinal sum M (mod 9) captures a latent variable influencing positional letter placement across the entire lexicon.

Several exact numerical matches between word sums and fundamental physical quantities were identified. These include M = 56 for IRON (mass number of the most stable iron isotope), W = 135 for PION (neutral pion mass), and M = 137 for AUTHORITY (inverse fine-structure constant), as well as the separation of degenerate linear cases (PION/SIFT) to within 0.22% and 0.02% of the respective constants via the cyclic operator. While individual matches could occur by chance in a large corpus, their clustering at physically significant values and the presence of semantically related terms warrant further targeted investigation.

The origin of the observed structure remains unresolved. Possible contributing factors include phonological constraints, etymological inheritance, morphological regularity, or selection pressures during language evolution that indirectly favour certain positional ordinal balances. The exact convergence on 13.5—the median of the alphabetic range 1–26—raises the additional question of whether the 26-letter system itself interacts with these constraints in a non-arbitrary manner.

No causal mechanism linking lexical statistics to physical constants is proposed here. The matches are reported as empirical observations requiring independent confirmation and deeper theoretical context.

Future work should (1) extend the analysis to other alphabetic languages and orthographies, (2) develop predictive models of w based on linguistic features, and (3) apply similar positional weighting schemes to non-linguistic sequences for comparative baseline strength.

The complete dataset and code are released under CC0 to facilitate replication and extension.

5. Conclusion

This analysis of 370,339 unique English words establishes strong statistical evidence for non-random positional ordering of letter ordinal values. The position-weighted mean ordinal value w converges exactly on 13.5 in digital root tier 9, forming a sharp peak with density exceeding all tested null models (within-word permutation, global shuffle, and bigram-generated pseudo-words) by more than 30 standard deviations ($p < 10^{-60}$). Distinct modal values of w appear across all nine digital root tiers, with highly significant separations.

Several precise numerical correspondences between word sums and fundamental physical quantities are observed, including:

- Simple ordinal sum M = 56 for IRON (mass number of ^{56}Fe, the most stable nuclide by binding energy per nucleon)
- Weighted positional sum W = 135 for PION (neutral pion rest mass 134.977 MeV/c^2)
- Simple ordinal sum M = 137 for AUTHORITY (inverse fine-structure constant ≈137.036)

These findings indicate systematic structure in the positional assignment of letter values within the English lexicon. The origin and implications of this structure remain open questions requiring further investigation across languages, historical stages, and alternative numbering schemes.

The complete dataset, analysis scripts, and reproducibility instructions are released under CC0 Universal (public domain) to enable independent verification and extension.

End of Document

Appendix A: Reproducibility Instructions

All analyses can be fully reproduced using the provided dataset Symphony135137.csv.Required environment: Python 3.12 with pandas, numpy, matplotlib, scipy.

Key regeneration commands:

```
# Install dependencies (if needed)

pip install pandas numpy matplotlib scipy


# Example script to regenerate Figure 1 (3×3 tier histograms) and Figure 2 (Tier 9 detail)

python - <<EOF

import pandas as pd

import matplotlib.pyplot as plt


df = pd.read_csv('Symphony135137.csv')

df['w'] = 2 * df['W'] / (df['n'] * (df['n'] + 1))

df['dr'] = 1 + (df['M'] - 1) % 9


# Figure 1: 3x3 grid

fig, axes = plt.subplots(3, 3, figsize=(15,15))

for tier in range(1,10):

    ax = axes[(tier-1)//3, (tier-1)%3]

    data = df[df['dr'] == tier]['w']

    ax.hist(data, bins=300, range=(3,25), density=True)

    ax.set_title(f'Tier {tier}')

plt.tight_layout()

plt.savefig('figure1_tiers.png')

plt.close()


# Figure 2: Tier 9 peak

tier9 = df[df['dr'] == 9]
```

```
plt.hist(tier9['w'], bins=1000, range=(13,14), color='blue', alpha=0.7,
label='Real')

plt.axvline(13.5, color='red', linestyle='--')

plt.title('Tier 9 Distribution Around w = 13.5')

plt.xlabel('w')

plt.ylabel('Count')

plt.legend()

plt.savefig('figure2_tier9_peak.png')

EOF
```

Full replication scripts are available upon request or can be reconstructed from the above template.Appendix B: Detailed Tier 9 Distribution

- Total words in Tier 9: 41,050
- Words with $|w - 13.5| < 0.001$: 2,887
- Peak bin density ($w = 13.495–13.505$): 1,915 counts per 0.01 bin
- Standard deviation of w in Tier 9: 0.42

Selected words at exact $w = 13.500$ (partial list; full sortable in CSV): LOVE, PLAY, ALWAYS, FULLNESS, VACUUM, BUSINESS, PION, SUN, SORROW, BY, PENS, STARS, NIGHT, THINKS, TRAINS, FUNNIEST, URAN…Appendix C: Corpus Statistics and Exact Match ListsExact word counts per digital root tier:

Tier	Word Count
1	41,306
2	41,265
3	41,054
4	41,080
5	40,916
6	41,140
7	41,361
8	41,215
9	41,050

Selected exact integer matches (full lists filterable in CSV):

- Simple ordinal sum M = 56: IRON, ATOM, STAR, BUCK, BALL, NOVA, BOND, GOOD, LIFE…
- Weighted positional sum W = 135: PION, SIFT (degenerate pair), …
- Simple ordinal sum M = 137: AUTHORITY, CONTROL, FEDERAL, COMPANY…
- Weighted positional sum W = 146: IRON, STEEL, MAGNET, BLADE, CORE…
- Weighted positional sum W = 183: LIGHT, PHOTON, VISION…
- Weighted positional sum W = 239: TRUTH, PLUTONIUM…

Appendix D: Cyclic Pairwise Product Operator and Spectral FingerprintFor a word with ordinal values L_1, L_2, …, L_n, the cyclic pairwise product sum is
C = ∑_{i=1}^{n} L_i · L_{(i mod n)+1}The 10-parameter spectral fingerprint used to break linear degeneracy combines:

1. Word length n
2. Simple sum M
3. Weighted sum W
4. Weighted mean w
5. Forward cyclic product C_forward
6. Reverse cyclic product C_reverse
   7–8. Mean and variance of absolute adjacent differences |L_{i+1} − L_i|
   9–10. M mod 9 and W mod 9

Step-by-step application to PION (ordinals 16,9,15,14):

- Linear metrics identical to SIFT
- Cyclic forward: 16·9 + 9·15 + 15·14 + 14·16 = 144 + 135 + 210 + 224 = 713
- Additional parameters yield final separated value 135.269 (dev. +0.22% from π° mass)

Step-by-step application to SIFT (19,9,6,20): yields 137.069 (dev. +0.02% from α^{-1}).

Full replication scripts are available upon request or can be reconstructed from the above template.

Appendix B: Detailed Tier 9 Distribution

- Total words in Tier 9: 41,050
- Words with $|w - 13.5| < 0.001$: 2,887
- Peak bin density ($w = 13.495–13.505$): 1,915 counts per 0.01 bin
- Standard deviation of w in Tier 9: 0.42

Selected words at exact $w = 13.500$ (partial list; full sortable in CSV):
LOVE, PLAY, ALWAYS, FULLNESS, VACUUM, BUSINESS, PION, SUN, SORROW, BY, PENS, STARS, NIGHT, THINKS, TRAINS, FUNNIEST, URAN…

Appendix C: Corpus Statistics and Exact Match Lists

Exact word counts per digital root tier:

Tier	Word Count
9	41,050
8	41,215
7	41,361
6	41,140
5	40,916
4	41,080
3	41,054
2	41,265
1	41,306

Selected exact integer matches (full lists filterable in CSV):

- Simple ordinal sum M = 56: IRON, ATOM, STAR, BUCK, BALL, NOVA, BOND, GOOD, LIFE…
- Weighted positional sum W = 135: PION, SIFT (degenerate pair), …
- Simple ordinal sum M = 137: AUTHORITY, CONTROL, FEDERAL, COMPANY…
- Weighted positional sum W = 146: IRON, STEEL, MAGNET, BLADE, CORE…
- Weighted positional sum W = 183: LIGHT, PHOTON, VISION…
- Weighted positional sum W = 239: TRUTH, PLUTONIUM…

Appendix D: Cyclic Pairwise Product Operator and Spectral Fingerprint

For a word with ordinal values $L_1, L_2, \ldots, L_n$, the cyclic pairwise product sum is
$C = \sum_{i=1}^{n} L_i \cdot L_{(i \bmod n)+1}$The 10-parameter spectral fingerprint used to break linear degeneracy combines:

1. Word length n
2. Simple sum M
3. Weighted sum W
4. Weighted mean w
5. Forward cyclic product C_forward
6. Reverse cyclic product C_reverse
   7–8. Mean and variance of absolute adjacent differences $|L_{i+1} - L_i|$
   9–10. M mod 9 and W mod 9

Step-by-step application to PION (ordinals 16,9,15,14):

- Linear metrics identical to SIFT
- Cyclic forward: $16 \cdot 9 + 9 \cdot 15 + 15 \cdot 14 + 14 \cdot 16 = 144 + 135 + 210 + 224 = 713$
- Additional parameters yield final separated value 135.269 (dev. +0.22% from π^0 mass)

Step-by-step application to SIFT (19,9,6,20): yields 137.069 (dev. +0.02% from α^{-1}).

End Of Appendices